



(19)
Bundesrepublik Deutschland
Deutsches Patent- und Markenamt

(10) **DE 601 20 949 T2** 2007.07.12

(12) **Übersetzung der europäischen Patentschrift**

(97) **EP 1 273 205 B1**

(51) Int Cl.⁸: **H04R 25/00** (2006.01)

(21) Deutsches Aktenzeichen: **601 20 949.4**

(86) PCT-Aktenzeichen: **PCT/DK01/00226**

(96) Europäisches Aktenzeichen: **01 919 235.0**

(87) PCT-Veröffentlichungs-Nr.: **WO 2001/076321**

(86) PCT-Anmeldetag: **04.04.2001**

(87) Veröffentlichungstag
der PCT-Anmeldung: **11.10.2001**

(97) Erstveröffentlichung durch das EPA: **08.01.2003**

(97) Veröffentlichungstag
der Patenterteilung beim EPA: **21.06.2006**

(47) Veröffentlichungstag im Patentblatt: **12.07.2007**

(30) Unionspriorität:
200000554 04.04.2000 DK

(84) Benannte Vertragsstaaten:
**AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT,
LI, LU, MC, NL, PT, SE, TR**

(73) Patentinhaber:
GN Resound A/S, Taastrup, DK

(72) Erfinder:
**NORDQVIST, Peter, Nils, S-191 42 Sollentuna, SE;
LEIJON, Arne, S-115 31 Stockholm, SE**

(74) Vertreter:
**Patentanwälte von Kreisler, Selting, Werner et col.,
50667 Köln**

(54) Bezeichnung: **EINE HÖRPROTHESE MIT AUTOMATISCHER HÖRUMGEBUNGSKLASSIFIZIERUNG**

Anmerkung: Innerhalb von neun Monaten nach der Bekanntmachung des Hinweises auf die Erteilung des europäischen Patents kann jedermann beim Europäischen Patentamt gegen das erteilte europäische Patent Einspruch einlegen. Der Einspruch ist schriftlich einzureichen und zu begründen. Er gilt erst als eingelegt, wenn die Einspruchsgebühr entrichtet worden ist (Art. 99 (1) Europäisches Patentübereinkommen).

Die Übersetzung ist gemäß Artikel II § 3 Abs. 1 IntPatÜG 1991 vom Patentinhaber eingereicht worden. Sie wurde vom Deutschen Patent- und Markenamt inhaltlich nicht geprüft.

Beschreibung

Gebiet der Erfindung

[0001] Die vorliegende Erfindung betrifft eine Hörprothese und ein Verfahren, das eine automatische Identifizierung oder Klassifizierung einer Hörumgebung bewirkt, indem es ein oder mehrere vorbestimmte Hidden-Markov-Modelle anwendet, um aus der Hörumgebung erhaltene akustische Signale zu verarbeiten. Die Hörprothese kann festgestellte Klassifizierungsergebnisse zum Steuern von Parameterwerten eines vorbestimmten Signalverarbeitungsalgorithmus oder zum Steuern des Umschaltens zwischen verschiedenen voreingestellten Hörprogrammen verwenden, um so die Signalverarbeitung der Hörprothese optimal an eine gegebene Hörumgebung anzupassen.

Hintergrund der Erfindung

[0002] Heutige digital gesteuerte oder Digitalsignalverarbeitung (DSP) verwendende Hörinstrumente sind häufig mit einer Anzahl voreingestellter Hörprogramme versehen. Die voreingestellten Hörprogramme sind oft vorgesehen, um eine komfortable und verständlich reproduzierte Klangqualität in verschiedenen Hörumgebungen zu erreichen. Aus dieser Hörumgebung erhaltene Audiosignale können höchst unterschiedliche Eigenschaften haben, beispielsweise bezüglich der durchschnittlichen und maximalen Schalldruckpegel (SPL) und/oder des Frequenzgehalts. Bei einer auf DSP basierenden Hörprothese kann daher jede Art von Hörumgebung eine besondere Einstellung von Algorithmusparametern eines Signalverarbeitungsalgorithmus der Hörprothese erfordern, um sicherzustellen, dass dem Benutzer in allen Arten von Hörumgebungen eine optimal reproduzierte Signalqualität zu Verfügung steht. Algorithmusparameter, die üblicherweise von einem Hörprogramm zum nächsten eingestellt werden können, umfassen Parameter bezüglich der Breitbandverstärkung, Eckfrequenzen oder Steigungen von frequenzselektiven Filteralgorithmen und Parameter, welche beispielsweise Knickstellenpunkte und Kompressionsverhältnisse von automatischen Verstärkungsregelungsalgorithmen (AGC) steuern. Infolgedessen sind die heutigen auf DSP basierenden Hörhilfen üblicherweise mit einer Anzahl verschiedener voreingestellter Hörprogramme versehen, die jeweils auf eine bestimmte Hörumgebung und/oder bestimmte Benutzerpräferenzen zugeschnitten sind. Die Eigenschaften der voreingestellten Hörprogramme werden üblicherweise während eines ersten Anpasstermins in den Räumlichkeiten eines Verkäufers festgelegt und in die Hilfe durch Übertragen oder Aktivieren entsprechender Algorithmen und Algorithmusparameter an bzw. in einem nicht-flüchtigen Speicherbereich der Hörprothese.

[0003] Dem Benutzer des Hörgeräts bleibt anschließend die Aufgabe überlassen, üblicherweise durch Betätigen eines Druckknopfes an der Hörhilfe oder eines Programmknopfes an einer Fernbedienung, manuell zwischen den voreingestellten Hörprogrammen entsprechend der aktuellen Hör- oder Klangumgebung zu wählen. Daher kann es vorkommen, dass der Benutzer, wenn er sich in den zahlreichen verschiedenen Klangumgebungen in seiner/ihrer täglichen Umwelt befindet und diese wieder verlässt, seine Aufmerksamkeit der gelieferten Klangqualität widmen und stetig die hinsichtlich der komfortablen Klangqualität und/oder der besten Sprachverständlichkeit beste Programmeinstellung suchen muss.

[0004] Es ist daher höchst erstrebenswert, eine Hörprothese, wie eine Hörhilfe oder eine Cochlea-Implantatvorrichtung, zu schaffen, die in der Lage ist, die gegenwärtige Hörumgebung eines Benutzers automatisch als zu einer von mehreren typischen alltäglichen Hörumgebungen gehörig zu klassifizieren. Anschließend können die Klassifizierungsergebnisse in der Hörprothese verwendet werden, um die Algorithmusparameter des gegenwärtigen Hörprogramms anzupassen, oder um zu einem anderen, geeigneteren voreingestellten Hörprogramm zu wechseln, um die optimale Klangqualität und/oder Sprachverständlichkeit für den individuellen Hörhilfenbenutzer aufrecht zu erhalten.

[0005] In der Vergangenheit wurden Versuche unternommen, die Signalverarbeitungscharakteristiken einer Hörhilfe an die Art der Hörumgebung anzupassen, in der sich der Benutzer befindet. US 5 687 241 beschreibt ein auf DSP basierendes Mehrkanal-Hörinstrument, das eine kontinuierliche Bestimmung oder Berechnung eines oder mehrerer Prozentwerte von Eingangssignalamplitudenverteilungen verwendet. Verstärkungswerte in den Frequenzkanälen werden anschließend in Reaktion auf die erkannten Pegel von Sprache und Geräuschen verändert. Jedoch ist es oft erstrebenswert, zwischen feinen Eigenschaften des Eingangssignals der Hörhilfe und nicht nur zwischen Sprache und Geräuschen zu unterscheiden. Zum Beispiel kann es erstrebenswert sein, zwischen einem Surround-Mikrofon- und einem Richtmikrofon-Hörprogramm nicht nur in Abhängigkeit von dem Hintergrundgeräuschpegel, sondern von weiteren Signaleigenschaften dieses Hintergrundgeräuschs zu wechseln. In Situationen, in denen der Benutzer der Hörprothese mit einer anderen Person bei vorhandenen Hintergrundgeräuschen kommuniziert, wäre es vorteilhaft, wenn es möglich wäre, die Art der Hin-

tergrundgeräusche identifizieren und klassifizieren zu können. Der Surround-Betrieb könnte in dem Fall gewählt werden, dass die Geräusche Verkehrsgeräusche sind, um es dem Benutzer zu ermöglichen, sich nähernden Verkehr unabhängig von der Annäherungsrichtung deutlich zu hören. Wenn andererseits die Hintergrundgeräusche als Störgeräusche klassifiziert werden, kann das Richt-Hörprogramm gewählt werden, um es dem Benutzer zu ermöglichen, ein reproduziertes Signal mit einem verbesserten Signal-Rauschverhältnis während einer Unterhaltung mit einer anderen Person zu erhalten.

[0006] Eine derartige detaillierte Charakterisierung eines Eingangssignals aus einer Hörumgebung kann durch Anwenden eines Hidden-Markov-Modells für die Analyse und Klassifizierung des Eingangssignals erreicht werden. Hidden-Markov-Modelle sind in der Lage, stochastische Eingangssignale hinsichtlich sowohl lang-, als auch kurzfristiger zeitlicher Variationen zu modellieren, anstatt auf das Modellieren langfristiger Amplitudenverteilungsstatistiken oder durchschnittlicher Energie beschränkt zu sein. Hidden-Markov-Modelle sind auf dem Gebiet der Spracherkennung als ein Mittel zum Modellieren statistischer Eigenschaften stochastischer Sprachsignale bekannt. Der Artikel "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", veröffentlicht in Proceedings of the IEEE, vol. 77, Nr. 2, Februar 1989, enthält eine umfassende Beschreibung der Anwendung von Hidden-Markov-Modellen auf Probleme bei der Spracherkennung.

[0007] Der Artikel "HMM-based speech enhancement using pitch period information in voiced speech segments", veröffentlicht in 1997 IEEE International symposium on circuits and systems, beschreibt ein Signalverarbeitungssystem zur Sprachverbesserung unter Verwendung eines Wiener-Filters, wobei Hidden-Markov-Modelle zur Steuerung der Parameter des Wiener-Filters verwendet werden, um Rauschen in einem Sprachsignal zu unterdrücken.

[0008] Die Anmelder haben jedoch zum ersten Mal diskrete Hidden-Markov-Modelle auf die Aufgabe des Klassifizierens der Hörumgebung einer Hörprothese angewandt, um eine automatische Anpassung eines oder mehrerer Parameter eines vorbestimmten Signalverarbeitungsalgorithmus zu erreichen, der in Verarbeitungseinrichtungen der Hörprothese in Abhängigkeit von diesen Klassifizierungsergebnissen ausgeführt wird.

Überblick über die Erfindung

[0009] Eine Aufgabe der Erfindung ist es, eine Hörprothese zu schaffen, die sich automatisch an eine umgebende Hörumgebung anpasst, indem sie einen oder mehrere Algorithmusparameter eines vorbestimmten Signalverarbeitungsalgorithmus steuert, um es einem Benutzer zu ermöglichen, automatisch verständliche und angenehm verstärkte Töne in einer Vielzahl verschiedener Klangumgebungen zu erhalten.

[0010] Es ist eine weitere Aufgabe der Erfindung, eine Hörprothese zu schaffen, die ein Eingangssignal kontinuierlich und automatisch als zu einer von mehreren alltäglichen Hörumgebungen gehörig klassifiziert und die Klassifizierungsergebnisse einer Verarbeitungseinrichtung anzeigt, um dieser die genannte Steuerung der Algorithmusparameter zu ermöglichen.

Beschreibung der Erfindung

[0011] Die Erfindung betrifft eine Hörprothese nach Anspruch 1.

[0012] Die Hörprothese kann ein Hörgerät oder eine Hörhilfe wie beispielsweise ein Behind-The-Ear-Hörhilfe (BTE), eine In-The-Ear-Hörhilfe (ITE) oder eine Completely-In-The-Canal-Hörhilfe (CIC) sein. Das von dem Mikrofon erzeugte Eingangssignal kann ein Analogsignal oder ein Digitalsignal im Multibitformat oder im Einbitformat sein, das von einem Mikrofonverstärker/-puffer bzw. einem integrierten A/D-Wandler erzeugt wird. Vorzugsweise wird das Eingangssignal an die Verarbeitungseinrichtung als digitales Eingangssignal geliefert. Falls das Mikrophonsignal in analoger Form geliefert wird, wird es daher von einem geeigneten Analog-Digital-Wandler (A/D-Wandler), der in einer integrierten Schaltung der Hörprothese enthalten sein kann, vorzugsweise in ein entsprechendes digitales Eingangssignal umgewandelt. Das Mikrophonsignal kann verschiedenen Signalverarbeitungsvorgängen, wie Verstärken und Bandbreitenbegrenzung, unterzogen werden, bevor es an den A/D-Wandler angelegt wird, und anschließend kann es anderen Vorgängen, wie Dezimierung, unterzogen werden, bevor das digitale Eingangssignal an die Verarbeitungseinrichtung angelegt wird.

[0013] Der Ausgangswandler, der das verarbeitete Ausgangssignal in ein akustisches oder elektrisches Signal oder solche Signale umwandelt, kann ein herkömmlicher Hörhilfenlautsprecher, der oft als "Empfänger" bezeichnet wird, oder ein anderer Schalldruckwandler sein, der ein wahrnehmbares akustisches Signal an den Benutzer der Hörprothese liefert. Der Ausgangswandler kann auch eine Anzahl von Elektroden aufweisen, die

betriebsmäßig mit dem Hörnerv oder den Hörnerven des Benutzers verbunden sind.

[0014] In der vorliegenden Beschreibung und den Ansprüchen bezeichnet der Begriff "vorbestimmter Signalverarbeitungsalgorithmus" jeglichen Verarbeitungsalgorithmus, der von der Verarbeitungseinrichtung der Hörprothese durchgeführt wird, welche das verarbeitete Ausgangssignal aus dem Eingangssignal erzeugt. Der "vorbestimmte Signalverarbeitungsalgorithmus kann daher eine Vielzahl von Subalgorithmen oder Subroutinen umfassen, die jeweils eine bestimmte untergeordnete Aufgabe in dem vorbestimmten Signalverarbeitungsalgorithmus durchführen. Beispielsweise kann der vorbestimmte Signalverarbeitungsalgorithmus unterschiedliche Signalverarbeitungssubroutinen, wie frequenzselektives Filtern, Einzel- oder Mehrkanalkompression, adaptive Rückkopplungsunterdrückung, Spracherkennung und Rauschverringern, etc., umfassen.

[0015] Ferner können mehrere verschiedene Auswahlen aus den vorgenannten Signalverarbeitungssubroutinen zusammengestellt werden, um zwei, drei oder mehr verschiedene voreingestellte Programme zu bilden, zwischen welchen der Benutzer je nach seiner Präferenz wählen kann.

[0016] Der vorbestimmte Signalverarbeitungsalgorithmus weist einen oder mehrere zusammenhängende Algorithmusparameter auf. Diese Algorithmusparameter können üblicherweise in eine Anzahl von kleineren Parametersätzen unterteilt werden, wobei jeder derartige Parameteralgorithmussatz, wie zuvor erläutert, einem bestimmten Teil des vorbestimmten Signalverarbeitungsalgorithmus oder einer bestimmten Subroutine zugeordnet ist. Diese Parametersätze steuern bestimmte Eigenschaften ihrer jeweiligen Subroutinen, wie beispielsweise Eckfrequenzen und Steigungen von Filtern, Kompressionsschwellenwerte und Verhältnisse von Kompressoralgorithmen, Adaptationsraten und Sondersignalcharakteristiken von adaptiven Rückkopplungsunterdrückungsalgorithmen, etc.

[0017] Werte der Algorithmusparameter werden vorzugsweise vorübergehend in einem flüchtigen Datenspeicherbereich der Verarbeitungseinrichtung, beispielsweise in einem Daten-RAM-Bereich, während der Ausführung des vorbestimmten Signalverarbeitungsalgorithmus gespeichert. Ausgangswerte der Algorithmusparameter sind in einem nicht-flüchtigen Speicherbereich, beispielsweise einem EEPROM/Flash-Speicherbereich oder einem batteriegestützten RAM-Speicherbereich, gespeichert, damit diese Algorithmusparameter bei Unterbrechungen der Energieversorgung, die üblicherweise durch das Entfernen oder Austauschen der Batterie der Hörhilfe oder durch Betätigen eines EIN/AUS-Schalters durch den Benutzer bewirkt werden, bewahrt werden.

[0018] Die Verarbeitungseinrichtung kann einen oder mehrere Prozessoren und dessen/deren zugehörige Speicherschaltungen umfassen. Der Prozessor kann durch einen Festkomma- oder Gleitkomma-Digitalsignalprozessor (DSP) mit Single- oder Dual-MAC-Architektur gebildet sein, der sowohl die in dem vorbestimmten Signalverarbeitungsalgorithmus erforderlichen Berechnungen, als auch eine Anzahl von sogenannten Haushaltsaufgaben durchführt, wie das Überwachen und Lesen von Werten von externen Interfacesignalen und Programmierports. Alternativ kann die Verarbeitungseinrichtung einen DSP aufweisen, der Zahlenverarbeitung, d.h. Multiplikation, Addition, Division etc., ausführt, während ein handelsüblicher oder gar eigener Mikroprozessorkern die Haushaltsaufgaben durchführt, die meist logische Operationen und Entscheidungsfindung beinhalten.

[0019] Der DSP kann vom software-programmierbaren Typ sein, der den vorbestimmten Signalverarbeitungsalgorithmus entsprechend Befehlen ausführt, die in einem zugehörigen Programm-RAM-Bereich gespeichert sind. Ein in der Verarbeitungseinrichtung integrierter Daten-RAM-Bereich kann Ausgangs- und Zwischenwerte der zugehörigen Algorithmusparameter und andere Datenvariablen während der Ausführung des vorbestimmten Signalverarbeitungsalgorithmus sowie verschiedene andere Haushaltsvariablen speichern. Ein derartiger software-programmierbarer DWSP kann für einige Anwendungen aufgrund der Möglichkeit des schnellen Implementierens und Testens von Modifikationen des vorbestimmten Signalverarbeitungsalgorithmus vorteilhaft sein. Die gleichen Vorteile gelten auch für Subroutinen, welche die Haushaltsaufgaben wahrnehmen. Alternativ kann die Verarbeitungseinheit durch einen festverdrahteten DSP-Kern gebildet drahteten DSP-Kern gebildet sein, so dass ein oder mehrere festgelegte vorbestimmte Signalverarbeitungsalgorithmen entsprechend einem festgelegten Befehlssatz eines zugeordneten Logikcontrollers ausgeführt werden. Bei dieser Art von festverdrahteter Prozessorarchitektur kann der Speicherbereich, der Werte der zugehörigen Algorithmusparameter speichert, in Form einer Registerdatei oder als RAM-Bereich ausgebildet sein, wenn die Anzahl der Algorithmusparameter die letztere Lösung rechtfertigt.

[0020] Erfindungsgemäß ist die Verarbeitungseinrichtung ferner in der Lage, das Eingangssignal in aufeinanderfolgende Signalblöcke der Dauer T_{frame} zu segmentieren und entsprechende Merkmalsvektoren $O(t)$ zu er-

zeugen, welche vorbestimmte Signalmerkmale der aufeinanderfolgenden Signalblöcke wiedergibt. Die Merkmalsvektoren werden anschließend mit mindestens einem Hidden-Markov-Modell, $\lambda^{\text{source}} = \{A^{\text{source}}, b(O(t)), \alpha_0^{\text{sq}}\}$, verarbeitet, das einer vorbestimmten Schallquelle zugeordnet ist, um Elementwerte eines Klassifizierungsvektors zu bestimmen. Dieser Klassifizierungsvektor gibt die Möglichkeit an, dass die vorbestimmte Schallquelle in der aktuellen Hörumgebung aktiv ist. Durch Regeln eines oder mehrerer Werte der dem vorbestimmten Signalverarbeitungsalgorithmus zugeordneten Algorithmusparameter in Abhängigkeit von Elementwerten des Klassifizierungsvektors wird die Verarbeitung des Eingangssignals an die Hörumgebung in Abhängigkeit von diesen Elementwerten angepasst. Die aufeinanderfolgenden Signalblöcke können nicht-überlappend oder mit einem vorbestimmten Überlappungsbetrag überlappend sein, d.h. zwischen 10% und 50% überlappend, um scharfe Diskontinuitäten an den Grenzen zwischen benachbarten Signalblöcken zu verhindern und/oder Fens-tereffekte einer beliebigen angewandten Fensterfunktion, beispielsweise ein Hanning-Fenster, an den Grenzen entgegenzuwirken. Während die genannte Blocksegmentierung des Eingangssignals zum Zweck des Erzeugens der Merkmalsvektoren $O(t)$ und zum Verarbeiten derselben mit dem mindestens einen Hidden-Markov-Modell erforderlich ist, kann der vorbestimmte Signalverarbeitungsalgorithmus das Eingangssignal Abtastung für Abtastung oder blockweise mit einer Blockzeit gleich oder verschieden von T_{frame} verarbeiten.

[0021] In den mehreren diskreten Hidden-Markov-Modellen $\lambda^{\text{source}} = \{A^{\text{source}}, B^{\text{source}}, \alpha_0^{\text{source}}\}$, bei denen B^{source} eine Beobachtungssymbolverteilungswahrscheinlichkeitsmatrix ist, welche als diskretes Äquivalent der generellen Funktion, $b(O(t))$, dient, welche die Wahrscheinlichkeitsfunktion für die Eingangsbeobachtung $O(t)$ für jeden Zustand eines Hidden-Markov-Modells definiert, ist die Verarbeitungseinrichtung vorzugsweise in der Lage, jeden der jeweiligen Merkmalsvektoren $O(t)$ mit einem oft als "Codebook" bezeichneten Merkmalsvektoren-satz zu vergleichen, um für im wesentlichen jeden der Merkmalsvektoren einen zugeordneten Symbolwert zu bestimmen, um eine Beobachtungsabfolge von den aufeinanderfolgenden Signalblöcken zugeordneten Symbolwerten zu erzeugen. Dieser Vorgang des Bestimmens von Symbolwerten aus den Merkmalsvektoren wird im allgemeinen als "Vektorquantisierung" bezeichnet. Anschließend wird die Beobachtungsabfolge von Symbolwerten mit dem mindestens einen Hidden-Markov-Modell λ^{source} verarbeitet, das der vorbestimmten Schallquelle zugeordnet ist, um die Elementwerte des Klassifizierungsvektors zu bestimmen.

[0022] Erfindungsgemäß ist die Verarbeitungseinrichtung in der Lage, die Merkmalsvektoren mit mehreren Hidden-Markov-Modellen zu verarbeiten oder die Beobachtungsabfolge von Symbolwerten mit mehreren diskreten Hidden-Markov-Modellen zu verarbeiten. Jedes der diskreten Hidden-Markov-Modelle oder jedes der Hidden-Markov-Modelle ist einer vorbestimmten Schallquelle zugeordnet, um die Elementwerte des Klassifizierungsvektors zu bestimmen. Jeder Elementwert kann direkt eine Wahrscheinlichkeit (d.h. einen Wert zwischen 0 und 1) der zugeordneten vorbestimmten Schallquelle wiedergeben, die in der gegenwärtigen Hörumgebung aktiv ist.

[0023] Die Dauer eines der Signalblöcke T_{frame} ist vorzugsweise so gewählt, dass sie innerhalb des Bereichs von 1–100 Millisekunden, beispielsweise ungefähr 5–10 Millisekunden, liegt. Eine derartige Zeitdauer ermöglicht das Operieren der Hidden-Markov-Modelle mit Zeitrahmen des Eingangssignals, die mit einzelnen Merkmalen, beispielsweise Phonemen, von Sprachsignalen vergleichbar sind, und mit Hüllkurvenmodulationen einer Anzahl von relevanten akustischen Geräuschquellen.

[0024] Eine vorbestimmte Schallquelle kann jede natürliche oder synthetische Schallquelle sein, wie eine natürliche Sprachquelle, eine Telefonsprachquelle, eine Verkehrsgeräuschquelle, eine Mehr-Sprecher- oder Störgeräuschquelle, eine U-Bahngeräuschquelle, eine Quelle vorübergehender Geräusche oder eine Windgeräuschquelle. Eine vorbestimmte Schallquelle kann auch durch eine Mischung aus natürlicher Sprache und/oder Verkehrsgeräuschen und/oder Störgeräuschen gebildet sein, die in vorbestimmten Verhältnissen gemischt sind, um beispielsweise ein bestimmtes Signalrauschverhältnis (snr) in dieser vorbestimmten Schallquelle zu erzeugen. Eine vorbestimmte Schallquelle kann Sprache und Hintergrundgeräusch sein, die in einem Verhältnis gemischt sind, das ein bestimmtes Ziel-Signalrauschverhältnis von beispielsweise 5 dB oder 10 dB oder vorzugsweise 20 dB erzeugt. Das einer derartigen gemischten Sprache-Hintergrundgeräuschquelle zugeordnete Hidden-Markov-Modell ist sodann durch den Klassifizierungsvektor in der Lage, anzugeben, wie gut ein gegenwärtiges Eingangssignal oder Eingangssignale zu dieser Sprache-Hintergrundgeräusch-Schallquelle passt bzw. passen. Die Verarbeitungseinrichtung kann infolgedessen geeignete Signalverarbeitungsparameter basierend auf sowohl dem Störgeräuschtypus, als auch das tatsächliche Signalrauschverhältnis wählen.

[0025] Zeit- und Spektralcharakteristiken jeder dieser vorbestimmten Schallquellen können auf der Basis von realen Aufzeichnungen von einer oder mehreren repräsentativen Schallquellen erhalten werden. Die Zeit- und Spektralcharakteristiken für jede Art von vorbestimmter Schallquelle werden vorzugsweise erhalten, indem eine reale Aufzeichnung einer Anzahl derartiger repräsentativer Schallquellen durchgeführt wird, und diese

Aufzeichnungen zu einer einzigen Aufnahme (oder Tondatei) verknüpft werden. Für Sprachschallquellen haben die Erfinder festgestellt, dass die Verwendung von ungefähr 10 verschiedenen Sprechern, vorzugsweise 5 männlichen und 5 weiblichen, im allgemeinen gute Klassifizierungsergebnisse in dem der Sprachquelle zugeordneten Hidden-Markov-Modell liefert. Der gemischte Schallquellentyp wird vorzugsweise gebildet, indem eine oder mehrere der realen Aufzeichnungen nachverarbeitet werden, um gewünschte spezifische Charakteristiken der gemischten Schallquelle zu erhalten, wie beispielsweise ein vorbestimmtes Signalrauschverhältnis.

[0026] Wenn die verknüpfte Schallquellenaufzeichnung gebildet wurde, werden Merkmalsvektoren, die vorzugsweise mit den Merkmalsvektoren identisch sind, die von der Verarbeitungseinrichtung in der Hörprothese erzeugt werden, aus der verknüpften Schallquellenaufzeichnung extrahiert, um eine Übungs-Beobachtungsabfolge für das zugeordnete kontinuierliche oder diskrete HMM zu erzeugen. Die Dauer der Übungsabfolge hängt von der Art der Schallquelle ab, jedoch hat sich gezeigt, dass eine Dauer von ungefähr 3–20 Minuten, beispielsweise ungefähr 4–6 Minuten, für zahlreiche Arten von Schallquellen einschließlich Sprachschallquellen adäquat ist. Anschließend wird für jede vorbestimmte Schallquelle das entsprechende HMM mit der erzeugten Übungsbeobachtungsabfolge trainiert, vorzugsweise mittels des iterativen Baum-Welch-Algorithmus, um Werte für A^{source} , die Zustandsübergangs-Wahrscheinlichkeitsmatrix, Werte für B^{source} , die Beobachtungssymbol-Wahrscheinlichkeitsmatrix (für diskrete HMM Modelle), und Werte für α_0^{source} , den Ausgangszustand-Wahrscheinlichkeitsverteilungsvektor, zu erhalten. Ist das HMM ergodisch, werden die Werte des Ausgangszustandswahrscheinlichkeitsverteilungsvektors aus der Zustandsübergangs-Wahrscheinlichkeitsmatrix bestimmt.

[0027] Die aus den aufeinanderfolgenden Signalblöcken erzeugten Merkmalsvektoren können Spektraleigenschaften der Signalblöcke, zeitliche Eigenschaften des Signalblocks oder eine Kombination derselben angeben. Die Spektraleigenschaften können in Form von diskreten Fourier-Transformationskoeffizienten, lineare prädiktive Kodierungsparameter, Cepstrum-Parameter oder entsprechender differentieller Cepstrum-Parameter vorliegen.

[0028] Wird ein diskretes HMM oder diskrete HMMs verwendet, kann das Codebook durch einen Offline-Übungsablauf bestimmt worden sein, der reale Schallquellenaufzeichnungen verwendete. Die Zahl der Merkmalsvektoren, welche das Codebook bilden, kann je nach der besonderen Anwendung variieren, jedoch hat sich für Hörhilfenanwendungen herausgestellt, dass ein Codebook mit zwischen 8 und 256 verschiedenen Merkmalsvektoren, wie beispielsweise 32–64 verschiedenen Merkmalsvektoren, üblicherweise eine adäquate Abdeckung des vollständigen Merkmalsraums bewirkt. Der Vergleich zwischen jedem der aus den aufeinanderfolgenden Merkmalsvektoren berechneten Merkmalsvektoren und dem Codebook liefert einen Symbolwert, der durch Wählen eines ganzzahligen Index wählbar ist, der zu dem Codebookeintrag gehört, der dem betreffenden Merkmalsvektor am nächsten ist. Das Ergebnis dieses Vektorquantisierungsprozesses kann eine Abfolge von ganzzahligen Indizes sein, welche die entsprechenden Symbolwerte repräsentieren.

[0029] Um das Codebook derart zu erzeugen, dass es den Merkmalsvektoren nahekommt, die in der Hörprothese während der Online-Verarbeitung des Eingangssignals, d.h. des normalen Gebrauchs, erzeugt werden, können die realen Tonaufzeichnungen durch Leiten des Signals durch einen Eingangssignalfad einer Ziel-Hörprothese erstellt worden sein. Durch Anwenden eines derartigen Vorgangs können Frequenzreaktionsabweichungen sowie andere lineare und/oder nichtlineare Verzerrungen, die durch den Eingangssignalfad der Ziel-Hörprothese erzeugt werden, durch Einbringen entsprechender Signalcharakteristiken in das Codebook kompensiert werden. Eine große Ähnlichkeit zwischen dem Merkmalsvektorsatz und online erzeugten Merkmalsvektoren wird sichergestellt, um die Erkennungs- und Klassifizierungsergebnisse der anschließenden Verarbeitung in dem diskreten Hidden-Markov-Modell oder Modellen zu optimieren. Ein ähnlicher vorteilhafter Effekt kann natürlich durch Durchführen einer Vorverarbeitung der realen Tonaufzeichnungen erhalten werden, die im wesentlichen der Verarbeitung des Eingangssignalfads einer Ziel-Hörprothese ähnlich ist, bevor der Merkmalsvektorsatz oder das Codebook extrahiert wird. Die letztere Lösung kann durch Anwenden geeigneter analoger und/oder digitaler Filter oder Filteralgorithmen auf das Eingangssignal implementiert werden, die zum Simulieren vorab bekannter Charakteristiken des betreffenden Eingangssignalfads ausgebildet sind.

[0030] Zwar hat es sich als nützlich erwiesen, sogenannte Von-links-nach-rechts-Hidden-Markov-Modelle auf dem Gebiet der Spracherkennung zu verwenden, bei denen die bekannten zeitlichen Charakteristiken von Wörtern und Äußerungen in einer Struktur des Modells Entsprechung finden, jedoch wurde von den Erfindern der vorliegenden Erfindung als vorteilhaft festgestellt, zumindest ein ergodisches Hidden-Markov-Modell zu verwenden, und vorzugsweise ergodische Hidden-Markov-Modelle für sämtliche angewendeten Hidden-Mar-

kov-Modelle einzusetzen. Ein ergodisches Hidden-Markov-Modell ist ein Modell, bei dem es möglich ist, jeden internen Zustand von einem anderen internen Zustand in dem Modell zu erreichen.

[0031] Die Anzahl der internen Modellzustände eines beliebigen bestimmten HMM der mehreren HMM kann von dem bestimmten Typ der nachgebildeten vorbestimmten Schallquelle abhängen. Eine relativ einfache, nahezu konstante Geräuschquelle kann durch ein HMM mit nur wenigen internen Zuständen adäquat nachgebildet werden, während komplexere Töne, wie Sprache oder gemischte Quellen von Sprache und komplexen Geräuschen, zusätzliche interne Zustände erfordern können. Vorzugsweise weist das mindestens eine Hidden-Markov-Modell oder jedes der mehreren Hidden-Markov-Modelle zwischen zwei und zehn Zustände auf, beispielsweise zwischen drei und acht Zustände. Gemäß einem bevorzugten Ausführungsbeispiel der Erfindung werden in einem Hörgerät vier diskrete HMM in einem eigenen DSP verwendet, wobei jedes der vier HMM vier interne Zustände hat. Die vier internen Zustände sind vier gewöhnlichen vorbestimmten Schallquellen zugeordnet: einer Sprachquelle, einer Verkehrsgeräuschquelle, einer Quelle mit mehreren Sprechern oder Störgeräuschen, und einer U-Bahngeräuschquelle. Ein Codebook mit 64 Merkmalsvektoren, die jeweils aus 12 Delta-Cepstrum-Parametern bestehen, bewirkt die Vektorquantisierung der aus dem Eingangssignal der Hörhilfe abgeleiteten Merkmalsvektoren. Jedoch kann der Merkmalsvektorsatz zwischen 8 und 256 verschiedene Merkmalsvektoren umfassen, beispielsweise 32–64 unterschiedliche Merkmalsvektoren, ohne übermäßigen Speicherplatz in dem unterschiedlichen Merkmalsvektoren, ohne übermäßigen Speicherplatz in dem DSP des Hörgeräts einzunehmen.

[0032] Die Verarbeitungseinrichtung kann in der Lage sein, das Eingangssignal entsprechend mindestens zweier unterschiedlicher vorbestimmter Signalverarbeitungsalgorithmen zu verarbeiten, von denen jeder einem Satz von Algorithmusparametern zugeordnet ist, wobei die Verarbeitungseinrichtung ferner in der Lage ist, einen Übergang zwischen den mindestens zwei vorbestimmten Signalverarbeitungsalgorithmen in Abhängigkeit von den Elementwerten des Klassifizierungsvektors zu steuern. Dieses Ausführungsbeispiel der Erfindung ist besonders nützlich, wenn die Hörprothese mit zwei eng beabstandeten Mikrofonen, beispielsweise zwei Surround-Mikrofonen, versehen ist, die zwei Eingangssignale erzeugen, welche zur Schaffung eines Richtsignalmodus durch bekannte Verzögerungssubtraktionsverfahren und eines ungerichteten Signalmodus beispielsweise durch Verarbeiten lediglich eines der Eingangssignale verwendet werden können. Die Verarbeitungseinrichtung kann den Übergang zwischen dem Richt- und dem Surround-Modus mittels eines Bereichs von Zwischenwerten der Algorithmusparameter stufenlos steuern, so dass die Richtionalität des Ausgangssignals allmählich zu-/abnimmt. Der Benutzer erfährt somit keine abrupten Änderungen in dem reproduzierten Ton, sondern vielmehr beispielsweise eine übergangslose Verbesserung des Signalrauschverhältnisses.

[0033] Um derartige Übergänge zwischen zwei vorbestimmten Signalverarbeitungsalgorithmen zu steuern, weist die Verarbeitungseinrichtung ferner eine Entscheidungssteuerungsvorrichtung auf, die die Elemente des Klassifizierungsvektors überwacht und Übergänge zwischen den mehreren Hidden-Markov-Modellen entsprechend einem vorbestimmten Regelsatz steuert. Die Entscheidungssteuerungsvorrichtung kann vorteilhafterweise als Zwischenschicht zwischen dem von den HMM gelieferten Klassifizierungsvektor und dem einen oder den mehreren von zugehörigen Algorithmusparametern arbeiten. Durch das Überwachen der Elementwerte des Klassifizierungsvektors und das Steuern des Werts bzw. der Werte des bzw. der zugehörigen Algorithmusparameter(s) nach Regeln bezüglich maximaler und minimaler Wechselzeiten zwischen HMM und, optional, Interpolationscharakteristiken zwischen den Algorithmusparametern, können die inhärenten Zeitrahmen, mit denen die HMM operieren, geglättet werden. Wenn beispielsweise eine Anzahl diskreter HMM mit aufeinanderfolgenden Symbolwerten operiert, die jeweils einen Zeitrahmen von ungefähr 6 ms darstellen, kann es vorteilhaft sein, schnelle Übergänge zwischen einem Sprach-HMM und einem Störgeräusch-HMM, die durch Pausen zwischen Wörtern in gesprochener Sprache in einer Hörumgebung vom Typ "Cocktail-Party" erzeugt werden, einer Tiefpassfilterung oder Glättung zu unterziehen. Anstelle eines sofortigen Wechsels zwischen den beiden vorbestimmten Signalverarbeitungsalgorithmen bei jedem Modellübergang, sind geeignete Zeitkonstanten und Hysterese in der Entscheidungssteuerungsvorrichtung vorgesehen.

[0034] Nach einem bevorzugten Ausführungsbeispiel der Erfindung weist die Entscheidungssteuerungsvorrichtung einen zweiten Satz von HMM auf, die mit einem wesentlich längeren Zeitrahmen des Eingangssignals operieren, als die HMM in einer ersten Schicht. Auf diese Weise ist die Verarbeitungseinrichtung in der Lage, die Beobachtungsabfolge von Symbolwerten oder die Merkmalswerte mit einem ersten Satz von Hidden-Markov-Modellen zu verarbeiten, die mit einem ersten Zeitrahmen operieren und einem ersten Satz von vorbestimmten Schallquellen zugeordnet sind, um Elementwerte eines ersten Klassifizierungsvektors zu bestimmen. Anschließend wird der erste Klassifizierungsvektor mit dem zweiten Satz von Hidden-Markov-Modellen verarbeitet, die mit einem zweiten Zeitrahmen operieren und einem zweiten Satz von vorbestimmten Schall-

quellen zugeordnet sind, um Elementwerte eines zweiten Klassifizierungsvektors zu bestimmen.

[0035] Der erste Zeitrahmen ist vorzugsweise innerhalb des Bereichs von 10–100 ms gewählt, um es dem ersten Satz von HMM zu ermöglichen, mit einzelnen Signalmerkmalen gewöhnlicher Sprach- und Geräuschsignale zu operieren, und der zweite Zeitrahmen ist vorzugsweise innerhalb des Bereichs von 1–60 Sekunden, beispielsweise ungefähr 10 bis 20 Sekunden, gewählt, um es dem zweiten Satz von HMM zu ermöglichen, mit Veränderungen zwischen verschiedenen Hörumgebungen zu operieren. Umgebungsbedingte Änderungen treten üblicherweise auf, wenn der Benutzer der Hörprothese sich zwischen unterschiedlichen Hörumgebungen bewegt, beispielsweise einer U-Bahnstation und dem Inneren eines Zuges oder einer häuslichen Umgebung, oder zwischen dem Inneren eines Autos und dem Stehen nahe einer Straße mit vorübergehendem Verkehr, etc.

Kurzbeschreibung der Zeichnungen

[0036] Ein bevorzugtes Ausführungsbeispiel einer erfindungsgemäßen Hörhilfe auf der Basis eines softwareprogrammierbaren DSP wird im folgenden unter Bezugnahme auf die Zeichnungen beschrieben, welche zeigen:

[0037] [Fig. 1](#) ein vereinfachtes Blockschaltbild einer erfindungsgemäßen Hörhilfe auf der Basis eines Drei-Chip-DSP, das Hidden-Markov-Modelle für die Eingangssignalklassifizierung verwendet,

[0038] [Fig. 2](#) ein Signalflussdiagramm eines vorbestimmten Signalverarbeitungsalgorithmus, der auf der Hörhilfe auf der Basis eines Drei-Chip-DSP von [Fig. 1](#) ausgeführt wird,

[0039] [Fig. 3](#) ein Signalflussdiagramm zur Darstellung eines Hörumgebungsklassifizierungsvorgangs,

[0040] [Fig. 4](#) ein Zustandsdiagramm für das Umgebungs-Hidden-Markov-Modell, das in [Fig. 3](#) als Block **550** dargestellt ist.

Detaillierte Beschreibung eines bevorzugten Ausführungsbeispiels

[0041] Im folgenden wird ein erfindungsgemäßes Ausführungsbeispiel einer Hörhilfe auf der Basis eines Drei-Chip-DSP detailliert beschrieben und erörtert. Die vorliegende Beschreibung erörtert im Detail nur die Operation des Signalverarbeitungsteils eines DSP-Kerns oder -Kernels mit den zugehörigen Speicherschaltungen. Die Gesamtschaltungstopologie, welche die Basis der DSP-Hörhilfe bilden kann, ist dem Fachmann bekannt und wird dementsprechend nur in sehr allgemeinen Begriffen dargelegt.

[0042] In dem vereinfachten Blockschaltbild der [Fig. 1](#) empfängt ein herkömmliches Hörhilfenmikrofon **105** ein akustisches Signal aus einer umgebenden Hörumgebung. Das Mikrofon **105** liefert ein analoges Eingangssignal an dem Anschluss MIC IN einer eigenen integrierten A/D-Schaltung **102**. Das analoge Eingangssignal wird in einem Mikrophonvorverstärker **106** verstärkt und an den Eingang eines ersten A/D-Wandlers einer Doppel-A/D-Wandlerschaltung **110** angelegt, die zwei synchron arbeitende Wandler vom Sigma-Delta-Typ aufweist. Ein serieller digitaler Datenstrom bzw. ein solches Signal wird in einer seriellen Interfaceschaltung **111** erzeugt und von dem Anschluss A/DDAT der eigenen integrierten A/D-Schaltung **102** an eine eigene Digitalsignalprozessorschaltung **2** (DSP-Schaltung) übertragen. Die DSP-Schaltung **2** weist einen A/D-Dezimator **13** auf, der in der Lage ist, den seriellen Digitaldatenstrom zu empfangen und ihn in entsprechende 16 Bit Audio-Samples mit einer geringeren Abtastrate zur weiteren Verarbeitung in einem DSP-Kern **5** umzuwandeln. Der DSP-Kern **5** hat einen zugehörigen Programm-Direktzugriffsspeicher (Programm-RAM) **6**, einen Daten-RAM **7** und einen Festwertspeicher (ROM) **8**. Die Signalverarbeitung des DSP-Kerns **5**, die im folgenden unter Bezugnahme auf das Signalflussdiagramm von [Fig. 2](#) beschrieben wird, wird durch Programmbefehle gesteuert, die aus dem Programm-RAM **6** gelesen werden.

[0043] Ein serielles bidirektionales Zweidraht-Programmierinterface **300** ermöglicht es einem (nicht dargestellten) Host-Programmiersystem mit der DSP-Schaltung **2** über eine serielle Interfaceschaltung **12** zu kommunizieren, und einem handelsüblichen EEPROM **202**, das Up-/Downloaden von Signalverarbeitungsalgorithmen und/oder zugehörigen Algorithmusparameterwerten.

[0044] Ein von dem DSP-Kern **5** aus dem analogen Eingangssignal erzeugtes digitales Eingangssignal wird an eine Impulsbreitenmodulationsschaltung **14** übertragen, welche empfangene Ausgangsabtastungen in ein impulsbreitenmoduliertes (PWM) und "noise-shaped"-verarbeitetes Ausgangssignal umwandelt. Das verarbei-

tete Ausgangssignal wird an zwei Anschlüsse des Hörhilfeempfängers **10** angelegt, der durch seine inhärente Tiefpassfiltereigenschaft das verarbeitete Ausgangssignal in ein entsprechendes akustisches Audiosignal umwandelt. Ein interner Taktgenerator und Verstärker **20** empfängt ein Master-Taktsignal von einer durch L1 und C5 gebildeten LC-Oszillator-Tankschaltung, welches zusammen mit einer internen Mastertaktschaltung **112** der A/D-Schaltung **102** einen Mastertakt sowohl für die DSP-Schaltung, als auch für die A/D-Schaltung **102** bildet. Der DSP-Kern **5** kann direkt durch das Mastertaktsignal oder durch ein geteiltes Taktsignal getaktet werden. Der DSP-Kern **5** ist vorzugsweise mit einer Frequenz von ungefähr 2–4 MHz getaktet.

[0045] **Fig. 2** zeigt eine relativ einfache Anwendung von Hidden-Markov-Modellen zum Steuern von Algorithmusparameterwerten eines vorbestimmten Signalverarbeitungsalgorithmus der DSP-basierten Hörhilfe von **Fig. 1**. Die diskreten Hidden-Markov-Modelle werden in der Hörhilfe oder dem Hörgerät verwendet, um eine automatische Klassifizierung dreier unterschiedlicher Hörumgebungen zu bewirken: Sprache in Verkehrsgereuschen, Sprache in Störgeräuschen, und reine Sprache, wie in **Fig. 4** dargestellt. Bei dem vorliegenden Ausführungsbeispiel der Erfindung ist jede Sprachumgebung mit einer bestimmten voreingestellten Frequenzantwort verbunden, die durch einen FIR-Filterblock **450** implementiert ist, der seine Filterparameterwerte von einer Filterwahlsteuerungsvorrichtung **430** empfängt. Die Operationen des FIR-Filterblocks **450** und der Filterwahlsteuerungsvorrichtung **430** erfolgen vorzugsweise durch jeweilige Subroutinen, die auf dem DSP-Kern **5** ausgeführt werden. Das Wechseln zwischen verschiedenen FIR-Filter-Parameterwerten erfolgt automatisch, wenn der Benutzer der Hörhilfe sich zwischen verschiedenen Hörumgebungen bewegt, was durch einen Hörumgebungsklassifizierungsalgorithmus **420** erkannt wird, der zwei Sätze diskreter HMM umfasst, die mit unterschiedlichen Zeitrahmen operieren, wie anhand der **Fig. 3** und **4** erläutert werden wird. Eine andere Möglichkeit besteht darin, den Hörumgebungsklassifizierer **420** einen zusätzlichen Mehrkanal-AGC-Algorithmus oder -System bereitstellen zu lassen, der zwischen den Eingang (IN) und den FIR-Filterblock **420** eingefügt werden könnte, und welcher Verstärkungswerte für aufeinanderfolgende Signalblöcke des Eingangssignals berechnet oder durch Tabellenreferenz bestimmt.

[0046] Der Benutzer kann eine bevorzugte Frequenzantwort/-verstärkung für jede der Hörumgebungen haben, die durch das entsprechende diskrete Hidden-Markov-Modell erkennbar/klassifizierbar ist. Diese bevorzugten Frequenzantworten/-verstärkungen können durch Anlegen einer Anzahl von vorgeschriebenen Standardverfahren wie NAL, POGO, etc. in Kombination mit individuellen interaktiven Feinabstimmungsverfahren.

[0047] In **Fig. 2** wird ein Roheingangssignal am Schaltungspunkt IN, das von dem Ausgang des A/D-Dezimators **13** in **Fig. 1** geliefert wird, zur Bildung aufeinanderfolgender Signalblöcke segmentiert, die jeweils eine Dauer von 6 ms haben. Das Eingangssignal wird vorzugsweise mit 16 kHz an diesem Schaltungspunkt abgetastet, so dass jeder Block aus 96 Audiosignalabtastungen besteht. Die Signalverarbeitung erfolgt über zwei unterschiedliche Wege: über einen Klassifizierungsweg durch Signalblöcke **410**, **420**, **440** und **430** und über einen vorbestimmten Signalverarbeitungsweg durch den Block **450**. Vorberechnete Impulsantworten der jeweiligen FIR-Filter werden während der Programmausführung in dem Daten-RAM gespeichert. Die Wahl der Parameterwerte oder -koeffizienten für den FIR-Filterblock **450** erfolgt durch den Filterauswahlblock **430** basierend auf den Elementwerten des Klassifizierungsvektors und, optional, auf den Daten des Spektrumschätzblocks **440**.

[0048] **Fig. 3** zeigt ein Signalflussdiagramm einer bevorzugte Ausbildung des Klassifizierungsblocks **420** von **Fig. 2**. Ein Vektorquantisiererblock (VQ) **510** geht der doppelagigen HMM-Architektur voran, wobei die Blöcke **520**, **521**, **522** eine erste HMM-Schicht bilden und der Block **550** eine zweite HMM-Schicht bildet. Das System besteht daher aus vier Stufen: einer Merkmalextraktionsschicht **500**, einer Schallmerkmalsklassifizierungsschicht **510**, der ersten HMM-Schicht in Form einer Schallquellenklassifizierungsschicht **520–522** und eine zweite HMM-Schicht in Form einer Hörumgebungsklassifizierungsschicht **550**. Die Schallquellenklassifizierungsschicht verwendet drei oder fünf Hidden-Markov-Modelle, und ein einzelnes HMM wird in der Hörumgebungsklassifizierungsschicht **550** verwendet.

[0049] Der Aufbau des Klassifizierungsblocks **420** ermöglicht es, unterschiedliche Wechselzeiten zwischen unterschiedlichen Hörumgebungen zu erreichen, beispielsweise langsamer Wechsel zwischen Verkehr und Hintergrund und schneller Wechsel zwischen Verkehr und Sprache.

[0050] Das Ausgangssignal OUT1 des Klassifizierungsblocks **420** ist ein Klassifizierungsvektor, bei dem jedes Element die Möglichkeit beinhaltet, dass eine bestimmte Schallquelle der drei vorbestimmten Schallquellen **520**, **521**, **522**, die durch ihre jeweiligen diskreten HMM nachempfunden sind, aktiv ist. Das Ausgangssignal OUT2 ist ein weiterer Klassifizierungsvektor, bei dem jedes Element die Wahrscheinlichkeit aufweist, dass eine bestimmte Hörumgebung aktiv ist.

[0051] Die Verarbeitung des Eingangssignals in dem vorgenannten Klassifizierungsweg wird im folgenden unter Bezugnahme auf die Implementierung in [Fig. 3](#) beschrieben.

[0052] Der Eingang zum Zeitpunkt t ist ein Block $x(t)$ der Größe B mit Eingangssignalabtastungen.

$$X(t) = [x_1(t), x_2(t) \dots x_B(t)]^T$$

[0053] $X(t)$ wird mit einem Fenster w_n multipliziert und die diskrete Fourier-Transformation DFT wird berechnet.

$$X_k(t) = \frac{1}{B} \sum_{n=0}^{B-1} w_n x_n(t) e^{-j \frac{2\pi kn}{B}} \quad k = 0..B/2-1$$

[0054] Ein Merkmalsvektor wird für jeden neuen Block extrahiert oder berechnet. gegenwärtig wird bevorzugt, 12 Cepstrum-Parameter für jeden Merkmalsvektor zu verwenden.

$$c_k(t) = \sum_{n=0}^{B/2-1} \cos\left(\frac{2\pi kn}{B}\right) \log|X_n(t)| \quad k = 0..11$$

[0055] Der Ausgang zum Zeitpunkt t ist ein Merkmalsspaltenvektor $f(t)$ mit durchgehenden Wertelementen.

$$f(t) = [c_0(t) \ c_1(t) \ \dots \ c_{11}(t)]^T$$

[0056] Der entsprechende differentielle Cepstrum-Parameterwert (oft als Delta-Cepstrum bezeichnet) wird als

$$\Delta f(t) = \sum_{i=0}^{K-1} h_i f(t-i),$$

wobei h_i derart bestimmt ist, dass $\Delta f(t)$ dem ersten Differential von $f(t)$ in bezug auf die Zeit t nahekommt. Die bevorzugte Länge des durch die Koeffizienten h_i definierten Filters ist $K = 8$.

[0057] Die Delta-Cepstrum-Koeffizienten werden dem Vektorquantisierer im Klassifizierungsblock **420** zugesandt. Andere Merkmale, beispielsweise Zeitdomänenmerkmale oder andere frequenzbasierte Merkmale können hinzugefügt werden.

[0058] Der Klassifizierungsblock **420** weist drei Schichten auf, die mit verschiedenen Zeitrahmen operieren: (1) eine Kurzzeitschicht (Schallmerkmalsklassifizierung) **510**, die sofort mit jedem Signalblock operiert, (2) eine Mittelfristschicht (Schallquellenklassifizierung) **501–522**, die im Zeitrahmen von Hüllkurvenmodulationen in vorbestimmten Schallquellen operiert, welche durch die vier HMM nachempfunden sind, und (3) eine Langzeitschicht (Hörumgebungsklassifizierung) **550**, die mit einem langsameren Zeitrahmen operiert, der Verschiebungen zwischen verschiedenen Schallquellen in einer bestimmten Hörumgebung oder der Verschiebung zwischen verschiedenen Hörumgebungen entspricht. Dies ist in [Fig. 4](#) dargestellt.

[0059] Die vorbestimmten Schallquellen, welche durch das vorliegende Ausführungsbeispiel der Erfindung nachempfunden werden, sind eine Verkehrsschallquelle, eine Hintergrundgeräuschquelle und eine reine Sprachquelle, könnten jedoch auch gemischte Schallquellen umfassen, die jeweils einen vorbestimmten Anteil von beispielsweise Sprache und Hintergrundgeräuschen oder Sprache und Verkehrsgeräuschen beinhalten, wie in [Fig. 4](#) dargestellt. Das letzte Ausgangssignal des Klassifizierers ist ein Hörumgebungswahrscheinlichkeitsvektor OUT1, der kontinuierlich eine aktuelle Wahrscheinlichkeitsschätzung für jede Hörumgebung umfasst, und ein Schallquellenwahrscheinlichkeitsvektor OUT2, der die geschätzte Wahrscheinlichkeit für jede Schallquelle angibt. Eine Hörumgebung kann aus einer der vorbestimmten Schallquellen **520–522** bestehen, oder aus einer Kombination von zwei oder mehr der vorbestimmten Schallquellen, wie in der Beschreibung der [Fig. 4](#) näher erläutert.

[0060] Der Eingang in den Vektorquantisiererblock **510** ist ein Merkmalsvektor mit Elementen mit kontinuierlichem Wert. Der Vektorquantisierer hat M , beispielsweise 32, Codewörter im Codebook $[c^1 \dots c^M]$, welche dem gesamten Merkmalsraum nahe kommen. Der Merkmalsvektor wird zu dem nächstliegenden Codewort in dem Codebook quantisiert und der Index $o(t)$, ein ganzzahliger Index zwischen 1 und M , des nächstliegenden Co-

deworts wird als Ausgang erzeugt.

$$O(t) = \arg \min_{i=1 \dots M} \left\| \Delta f(t) - c^i \right\|^2$$

[0061] Der VQ wird offline mit dem Generalized-Lloyd-Algorithmus trainiert (Linde, 1980). Das Trainingsmaterial umfasst reale Aufzeichnungen von Schallquellenabtastungen. Diese Aufzeichnungen wurden über den in [Fig. 1](#) dargestellten Eingangssignalfeld des DSP-basierten Hörgeräts erstellt.

[0062] Jede der drei Schallquellen ist durch ein diskretes HMM nachempfunden. Jedes HMM besteht aus einer Zustandsübergangs-Wahrscheinlichkeitsmatrix A^{source} , einer Beobachtungssymbolwahrscheinlichkeitsverteilungsmatrix B^{source} und einem Ausgangszustand-Wahrscheinlichkeitsvektor α_0^{source} .

[0063] Eine kompakte Notation für ein HMM ist $\lambda^{\text{source}} = \{A^{\text{source}}, B^{\text{source}}, \alpha_0^{\text{source}}\}$. Jedes Schallquellenmodell hat $N = 4$ interne Zustände und beobachtet den Strom von VQ-Symbolwerten oder Schwerpunktindizes $[O(1) \dots O(t)]$ $O_t \in [1, M]$. Der aktuelle Zustand zum Zeitpunkt t ist als stochastische Variable $Q_{\text{source}}(t) \in \{1, \dots, N\}$ modelliert.

[0064] Der Zweck der Mittelfristschicht ist es, zu schätzen wie gut jedes Quellenmodell die aktuelle Eingangsbeobachtung $O(t)$ erklären kann. Der Ausgang ist ein Spaltenvektor $u(t)$ mit Elementen, welche die bedingten Wahrscheinlichkeiten $\phi^{\text{source}}(t) = \text{prob}(O(t)|O(t-1), \dots, O(1), \lambda^{\text{source}})$ für jede Quelle angeben.

[0065] Der Standard-Vorwärtsalgorithmus (Rabiner, 1988) wird zum rekursiven Aktualisieren des Zustandswahrscheinlichkeitsspaltenvektors $p^{\text{source}(i)}$ verwendet. Die Elemente $p^{\text{source}(i)}(t)$ dieses Vektors geben die konditionale Wahrscheinlichkeit an, dass die Schallquelle sich im Zustand i befindet,

$$p_i^{\text{source}}(t) = \text{prob}(Q^{\text{source}}(t) = i, o(t)|o(t-1), \dots, o(1), \lambda^{\text{source}}).$$

[0066] Die Gleichungen der rekursiven Aktualisierung sind:

$$p_i^{\text{source}}(t) = ((A^{\text{source}})^T \hat{p}^{\text{source}}(t-1)) \cdot b^{\text{source}}(o(t))$$

$$\phi^{\text{source}}(t) = \text{prob}(o(t)|o(t-1), \dots, o(1), \lambda^{\text{source}}) = \sum_{i=1}^N p_i^{\text{source}}(t)$$

$$\hat{p}_i^{\text{source}}(t) = p_i^{\text{source}}(t) / \sum_{i=1}^N p_i^{\text{source}}(t)$$

wobei der Operator \otimes eine elementweise Multiplikation definiert.

[0067] [Fig. 4](#) zeigt detaillierter eine leicht modifizierte Version des zweilagigen HMM-Aufbaus von [Fig. 3](#), bei dem die erste Schicht von HMM **520–522** zwei zusätzliche HMM aufweist, nämlich ein viertes HMM, das eine vorbestimmte Schallquelle "Sprache in Verkehrsgerauschen" nachempfunden, und ein fünftes HMM, das eine vorbestimmte Schallquelle "Sprache in Cafeteria-Hintergrundgerauschen" nachempfunden.

[0068] Das Signal OUT1 der letzten HMM-Schicht **550** schätzt aktuelle Wahrscheinlichkeiten für jede der nachempfundenen Hörumgebungen durch Beobachten des von der vorherigen HMM-Schicht kommenden Stroms von Schallquellenwahrscheinlichkeitsvektoren. Die Hörumgebung ist durch eine diskrete stochastische Variable $E(t) \in \{1 \dots 3\}$ wiedergegeben, wobei die Ergebnisse codiert sind als 1 für "Sprache in Verkehrsgerauschen", 2 für "Sprache in Cafeteria-Hintergrundgerauschen" und 3 für "reine Sprache". Der Ausgangswahrscheinlichkeitsvektor oder Klassifizierungsvektor hat somit drei Elemente, nämlich eines für jede dieser Umgebungen. Die letzte HMM-Schicht **550** enthält fünf Zustände, die Verkehrsgerauschen, Sprache (in Verkehrsgerauschen, "Sprache/T"), Hintergrundgerauschen, Sprache (in Hintergrundgerauschen, "Sprache/B") und reine Sprache ("Sprache/C") wiedergeben. Übergänge zwischen Hörumgebungen, die durch gestrichelte Pfeile dargestellt sind, haben eine geringe Wahrscheinlichkeit und Übergänge zwischen Zuständen innerhalb einer Hörumgebung, die durch durchgezogene Pfeile dargestellt sind, haben relativ hohe Wahrscheinlichkeiten.

[0069] Die letzte HMM-Schicht **550** besteht aus einem Hidden-Markov-Modell mit fünf Zuständen und einer Übergangswahrscheinlichkeitsmatrix A^{env} ([Fig. 4](#)). Der aktuelle Zustand in dem Umgebungs-Hidden-Mar-

kov-Modell ist als eine diskrete stochastische Variable $S(t) \in \{1 \dots 5\}$ nachempfunden, wobei die Ergebnisse codiert sind als 1 für "Verkehr", 2 für Sprache (in Verkehrsgeräuschen, "Sprache/T"), 3 für Hintergrundgeräusche, 4 für Sprache (in Hintergrundgeräuschen, "Sprache/B") und 5 für reine Sprache ("Sprache/C").

[0070] Die Hörumgebung "Sprache in Verkehrsgeräuschen" $E(t) = 1$ hat zwei Zustände $S(t) = 1$ und $S(t) = 2$. Die Hörsituation "Sprache in Cafeteria-Hintergrundgeräuschen" $E(t) = 2$ hat zwei Zustände $S(t) = 3$ und $S(t) = 4$. Die Hörumgebung mit reiner Sprache $E(t) = 3$ hat nur einen Zustand $S(t) = 5$. Die Wahrscheinlichkeit der Übergänge zwischen Hörumgebungen sind relativ gering und die Wahrscheinlichkeit von Übergängen zwischen Zuständen innerhalb einer Hörumgebung sind hoch.

[0071] Das Umgebungs-Hidden-Markov-Modell **550** beobachtet den Strom der Vektoren $[u(1) \dots u(t)]$, wobei

$$u(t) = [\phi^{\text{traffic}}(t)\phi^{\text{speech}}(t)\phi^{\text{babble}}(t)\phi^{\text{speech}}(t)\phi^{\text{speech}}(t)]^T$$

die geschätzten Beobachtungswahrscheinlichkeiten für jeden Zustand enthält.

[0072] Die Wahrscheinlichkeit $\hat{p}_i^{\text{env}} = \text{prob}(S(t) = i | u(t), \dots, u(1), A^{\text{env}})$, unter Berücksichtigung der aktuellen und sämtlicher vorhergehender Beobachtungen und in Anbetracht des Umgebungs-Hidden-Markov-Modells, in einem bestimmten Zustand zu sein, wird durch den Vorwärtsalgorithmus

$$p^{\text{env}}(t) = ((A^{\text{env}})^T \hat{p}^{\text{env}}(t-1) \cdot u(t),$$

berechnet (Rabiner, 1989), mit Elementen

$$p_i^{\text{env}} = \text{prob}(S(t) = i, u(t) | u(t-1), \dots, u(1), A^{\text{env}}),$$

und schließlich, mit Normalisierung,

$$\hat{p}^{\text{env}}(t) = p^{\text{env}}(t) / \sum p_i^{\text{env}}(t).$$

[0073] Die Wahrscheinlichkeit für jede Hörumgebung $p^E(t)$ kann nunmehr, unter Berücksichtigung sämtlicher vorhergehender Beobachtungen und des Umgebungs-Hidden-Markov-Modells, berechnet werden als

$$\mathbf{p}^E(t) = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \hat{\mathbf{p}}^{\text{env}}(t).$$

[0074] Wie zuvor erwähnt, ist der Spektrumschätzungsblock **440** der [Fig. 2](#) optional, kann jedoch zum Schätzen eines durchschnittlichen Frequenzspektrums verwendet werden, das sich langsam an die aktuelle Hörumgebung anpasst. Eine andere Möglichkeit ist es, zwei oder mehr sich langsam anpassende Spektren für unterschiedliche Schallquellen in einer gegebenen Hörumgebung, beispielsweise ein Sprachspektrum und ein Geräuschkenspektrum, zu schätzen.

[0075] Die Geräuschwahrscheinlichkeiten $\phi^{\text{source}}(t)$, die Umgebungswahrscheinlichkeiten $p^E(t)$ und das aktuelle Log-Power-Spektrum $X(t)$ werden zum Schätzen der aktuellen Signal- und Rausch-Log-Power-Spektren verwendet. Bei der Schätzung werden zwei Tiefpassfilter eingesetzt, ein Filter für das Signalspektrum und ein Filter für das Rauschspektrum. Das Signalspektrum wird aktualisiert, wenn $p_1^E(t) > p_2^E(t)$ und $\phi^{\text{speech}}(t) > \phi^{\text{traffic}}(t)$ oder wenn $p_2^E(t) > p_1^E(t)$ und $\phi^{\text{speech}}(t) > \phi^{\text{babble}}(t)$. Das Rauschspektrum wird aktualisiert, wenn $p_1^E(t) > p_2^E(t)$ und $\phi^{\text{traffic}}(t) > \phi^{\text{speech}}(t)$ oder wenn $p_2^E(t) > p_1^E(t)$ und $\phi^{\text{babble}}(t) > \phi^{\text{speech}}(t)$.

Notation:

M Anzahl von Schwerpunkten im Vektorquantisierer

N Anzahl von Zuständen im HMM

$\lambda^{\text{source}} = \{A^{\text{source}}, B^{\text{source}}, \pi^{\text{source}}\}$ kompakte Notation für ein diskretes HMM, das eine Quelle mit N Zuständen und

M Beobachtungssymbolen beschreibt

B Blockgröße

O = $[O_{-\infty} \dots O_t]$ Beobachtungsabfolge

$O_t \in [1, M]$ diskrete Beobachtung zum Zeitpunkt t

$f(t)$ Merkmalsvektor

w Fenster der Größe B

$x(t)$ ein Block der Größe B zum Zeitpunkt t aus Rohreingangsmustern

$X(t)$ das entsprechende diskrete komplexe Spektrum mit der Größe B zum Zeitpunkt t

Literatur

L. R. Rabiner, A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. Proc. IEEE, vol. 77, no. 2, February 1989

Linde, Y., Buzo, A., and Gray, R. M. An Algorithm for Vector Quantizer Design. IEEE Trans. Comm., COM-28: 84–95, January 1980.

Europäisches Patent 0001273205

Deutsches Aktenzeichen 60120949.4

GN ReSound A/S

Patentansprüche

1. Hörprothese mit:

einem Mikrophon (**105**), das in der Lage ist, ein Eingangssignal in Reaktion auf den Empfang eines akustischen Signals aus einer Hörumgebung zu erzeugen,

einem Ausgangswandler (**10**) zum Umwandeln eines verarbeiteten Ausgangssignals in ein elektrisches oder akustisches Ausgangssignal,

einer Verarbeitungseinrichtung (**2**), die in der Lage ist, das Eingangssignal entsprechend einem vorbestimmten Signalverarbeitungsalgorithmus und zugehörigen Algorithmusparametern zu verarbeiten, um das verarbeitete Ausgangssignal zu erzeugen,

einem Speicherbereich (**202**), der Werte der zugehörigen Algorithmusparameter für den vorbestimmten Signalverarbeitungsalgorithmus speichert,

wobei die Verarbeitungseinrichtung (**2**) ferner in der Lage ist,

das Eingangssignal in konsekutive Zeitdauer-Signalblöcke T_{frame} zu segmentieren und jeweilige Merkmalsvektoren $O(t)$ zu erzeugen, welche die vorbestimmten Signalmerkmale der konsekutiven Signalblöcke wiedergeben,

dadurch gekennzeichnet, dass die Verarbeitungseinrichtung (**2**) ferner geeignet ist,

jeden der Merkmalsvektoren $O(t)$ mit einer Merkmalsvektorgruppe zu vergleichen, um im wesentlichen für jeden Merkmalsvektor einen zugehörigen Symbolwert zu bestimmen, um eine Beobachtungsabfolge von zu den konsekutiven Signalblöcken gehörenden Symbolwerten zu erzeugen;

die Beobachtungsabfolge von Symbolwerten durch mehrere diskrete Hidden-Markov-Modelle (**520, 521, 522**), $\lambda^{\text{source}} = \{A^{\text{source}}, B^{\text{source}}, c_0^{\text{source}}\}$, zu verarbeiten, die auf einer ersten Zeitskala arbeiten und vorbestimmten Schallquellen zugeordnet sind, um Elementwerte eines ersten Klassifizierungsvektors zu bestimmen, welcher die Möglichkeit angibt, dass die vorbestimmte Schallquelle in der Hörumgebung aktiv ist;

einen oder mehrere Werte der zugehörigen Algorithmusparameter in Abhängigkeit von dem/den Elementwert/-en des ersten Klassifizierungsvektors zu regeln;

wodurch Eigenschaften des vorbestimmten Signalverarbeitungsalgorithmus an gegenwärtige Schallquellen angepasst werden, die in der Hörumgebung aktiv sind, und

ferner mit einer Entscheidungssteuerungsvorrichtung, die in der Lage ist, die Elemente des ersten Klassifizierungsvektors zu überwachen und Übergänge zwischen Signalverarbeitungsalgorithmen entsprechend einem vorbestimmten Regelsatz zu steuern, der geeignete Zeitkonstanten und Hysterese zu schaffen, wobei

A^{source} = eine Zustandsübergangs-Wahrscheinlichkeitsmatrix bezeichnet,

B^{source} = eine Beobachtungssymbol-Wahrscheinlichkeitsverteilungsmatrix für eine Eingangsbeobachtung für jeden Zustand des mindestens einen Hidden-Markov-Modells bezeichnet,

c_0^{source} = einen Ausgangszustand-Wahrscheinlichkeitsverteilungsvektor bezeichnet.

2. Hörprothese nach Anspruch 1, bei der die Merkmalsvektoren jeweiligen ganzzahligen Symbolwerten während eines Vektorquantisierungsvorgangs zugeordnet werden.

3. Hörprothese nach einem der vorhergehenden Ansprüche, bei der die Merkmalsvektorgruppe zwischen 8 und 256 verschiedene Symbole umfasst.

4. Hörprothese nach einem der Ansprüche 1–3, bei der die Merkmalsvektorgruppe in einem Offline-Trainingsvorgang bestimmt wurde, welcher reale Schallquellenaufzeichnungen verwendete, die in nicht flüchtigen

Speicherstellen des Hörgeräts gespeichert sind.

5. Hörprothese nach Anspruch 4, bei der die realen Tonaufnahmen über einen Eingangspfad einer Ziel-Hörprothese oder durch Durchführen einer im wesentlichen ähnlichen Signalverarbeitung eines Eingangssignals zum Simulieren von Eigenschaften des Eingangssignalfads erfolgten.

6. Hörprothese nach einem der vorhergehenden Ansprüche, bei der die Entscheidungssteuervorrichtung (550) ein diskretes Hidden-Markov-Modell (550) aufweist, das über eine Zeitskala des Eingangssignals arbeitet, die wesentlich länger ist als die inhärenten Zeitskalen der mehreren diskreten Hidden-Markov-Modelle (520, 521, 522).

7. Hörprothese nach Anspruch 6, bei der die inhärenten Zeitskalen der mehreren diskreten Hidden-Markov-Modelle (520, 521, 522) innerhalb eines Bereichs zwischen 10–100 ms gewählt sind, und die wesentlich längere Zeitskala des diskreten Hidden-Markov-Modells (550) in einem Bereich von 1–60 Sekunden gewählt ist.

8. Hörprothese nach einem der Ansprüche 1–7, bei der die Entscheidungssteuervorrichtung ferner in der Lage ist, den ersten Klassifizierungsvektor mit einer zweiten Gruppe diskreter Hidden-Markov-Modelle (550) zu verarbeiten, die mit einer zweiten Zeitskala arbeiten und einer Gruppe von vorbestimmten Hörumgebungen zugeordnet sind, um Elementwerte eines zweiten Klassifizierungsvektors zu bestimmen, einen oder mehrere Werte der zugehörigen Algorithmusparameter in Abhängigkeit von Elementwerten des zweiten Klassifizierungsvektors zu regeln, wodurch Eigenschaften des vorbestimmten Signalverarbeitungsalgorithmus an eine gegenwärtige Hörumgebung angepasst werden.

9. Hörprothese nach einem der vorhergehenden Ansprüche, bei der der Wert von T_{frame} zwischen 1 und 100 Millisekunden liegt, beispielsweise zwischen ungefähr 5–10 Millisekunden.

10. Hörprothese nach Anspruch 8, bei der die erste Zeitskala innerhalb des Bereichs von 10–100 ms und die zweite Zeitskala innerhalb des Bereichs von 1–60 Sekunden gewählt ist.

11. Hörprothese nach einem der vorhergehenden Ansprüche, bei der das diskrete Hidden-Markov-Modell oder die diskreten Hidden-Markov-Modelle (520, 521, 522, 550) mindestens ein ergodisches diskretes Hidden-Markov-Modell umfasst/umfassen.

12. Hörprothese nach einem der vorhergehenden Ansprüche, bei der das mindestens eine vorbestimmte diskrete Hidden-Markov-Modell (520, 521, 522, 550) oder jedes der mehreren vorbestimmten diskreten Hidden-Markov-Modelle (520, 521, 522, 550) zwischen zwei und zehn Zustände aufweist.

Es folgen 3 Blatt Zeichnungen

Anhängende Zeichnungen

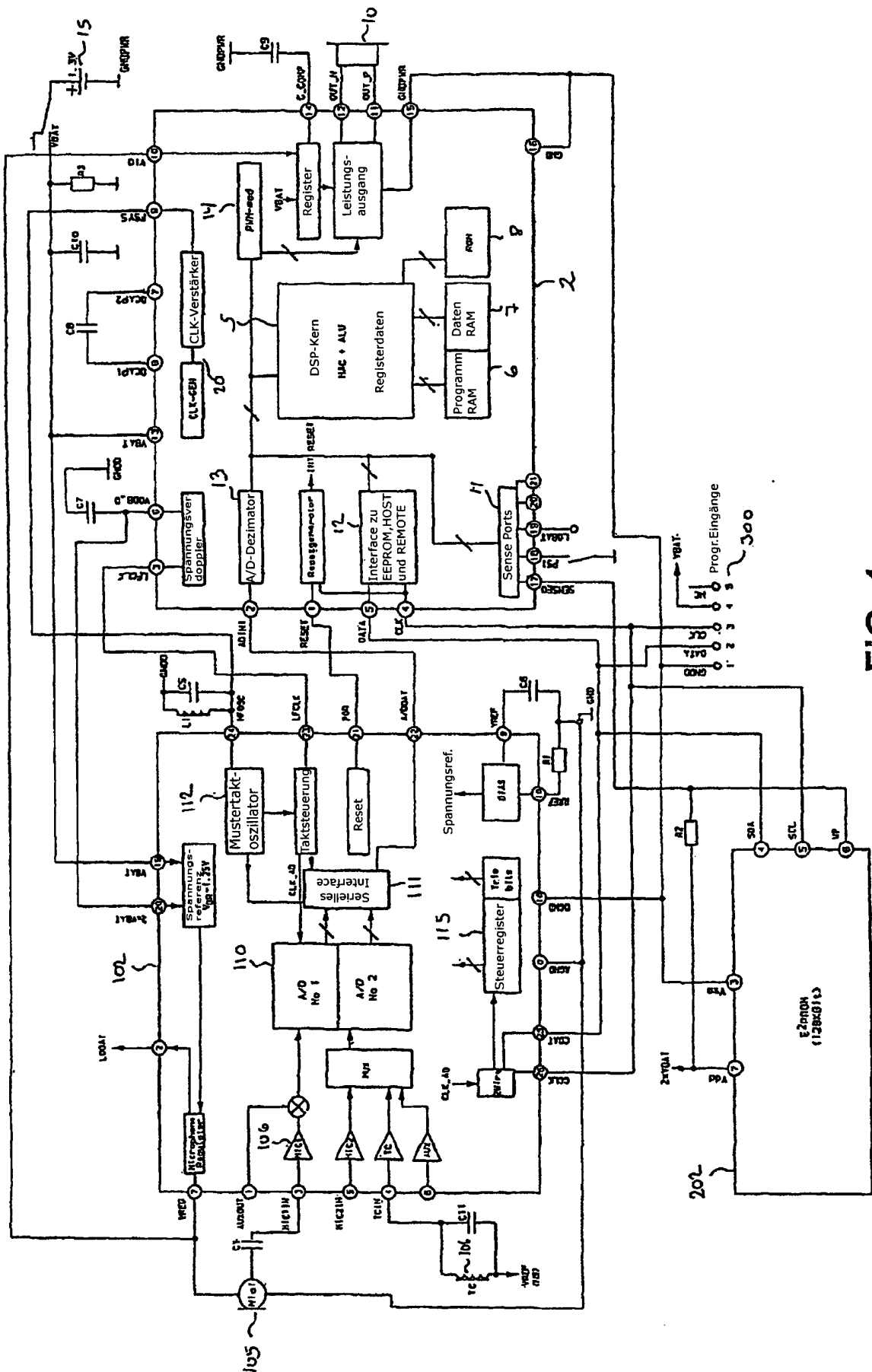


FIG. 1

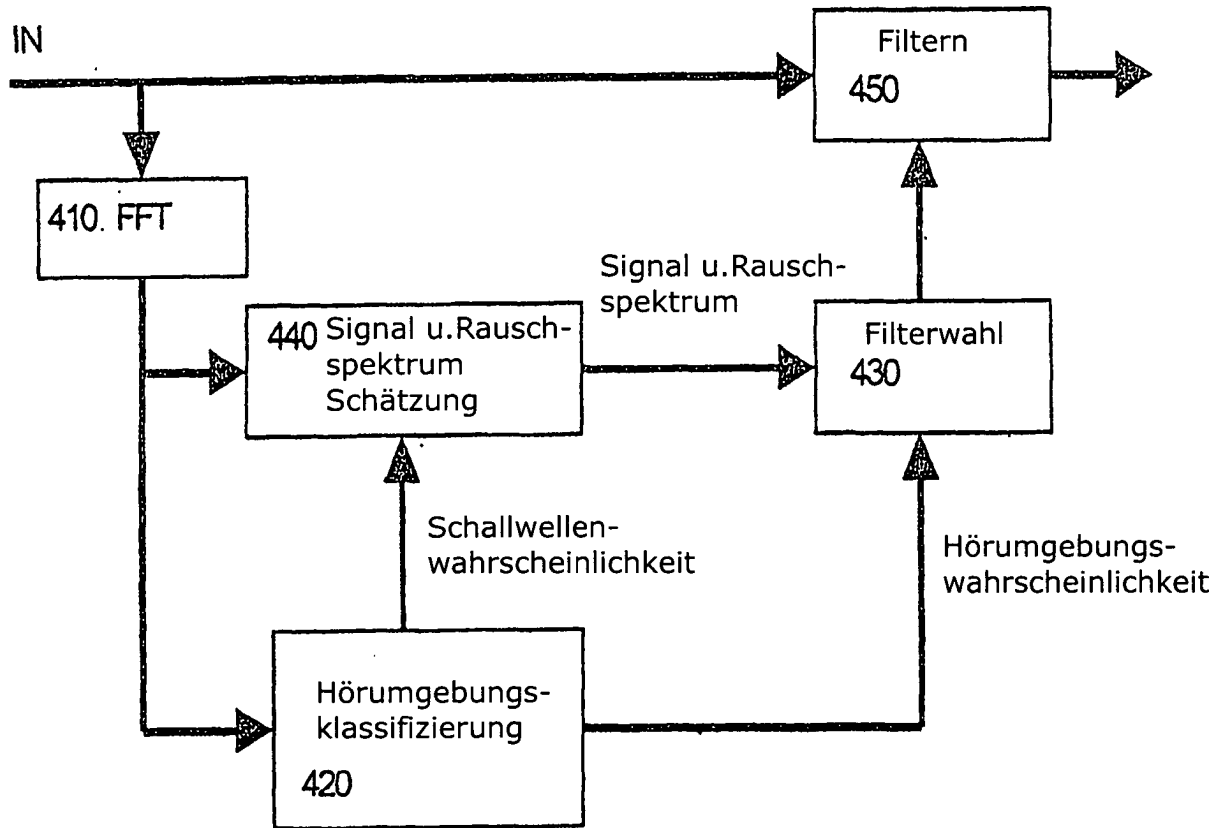


FIG. 2

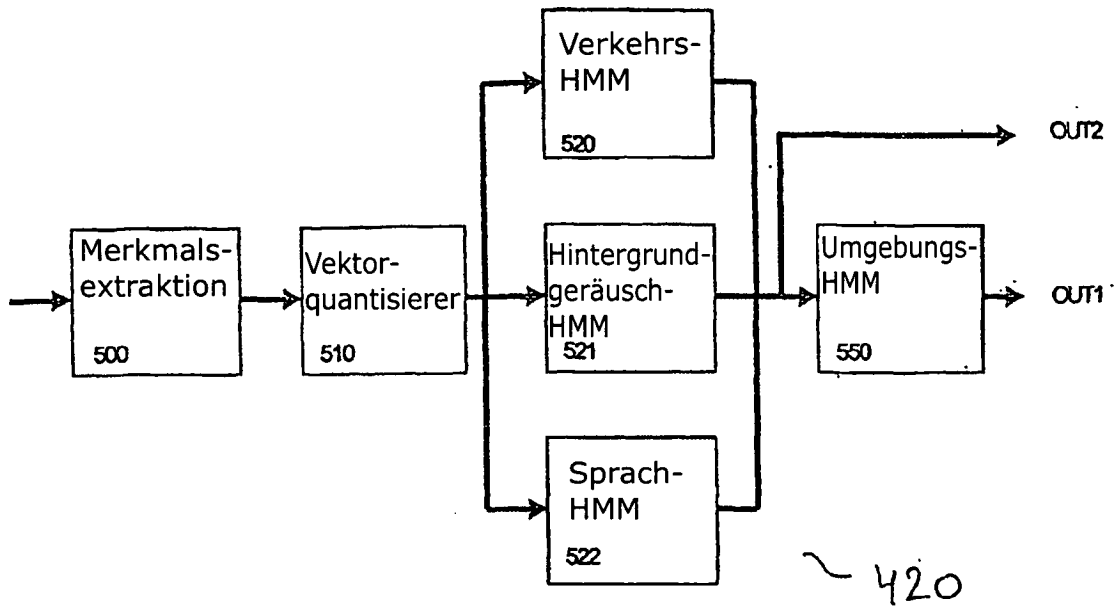


FIG. 3

